

◆ D24-075 ◆

應用於大規模自動導航最佳化之量子啟發式數位退火處理器晶片

A Fully Integrated Annealing Processor for Large-Scale Autonomous Navigation Optimization

隊伍名稱 | CVSD雷助教 CVSD Best TA

隊長 | 朱怡蓁 / 臺灣大學電子工程學研究所

隊員 | 林祐丞 / 臺灣大學電子工程學研究所

◆ 指導教授 ◆



楊家驥 | 臺灣大學電機工程學系

美國加州大學洛杉磯分校電機博士，現為臺灣大學電機工程學系暨電子工程學研究所教授。曾於交通大學電子工程學系任教。

研究領域

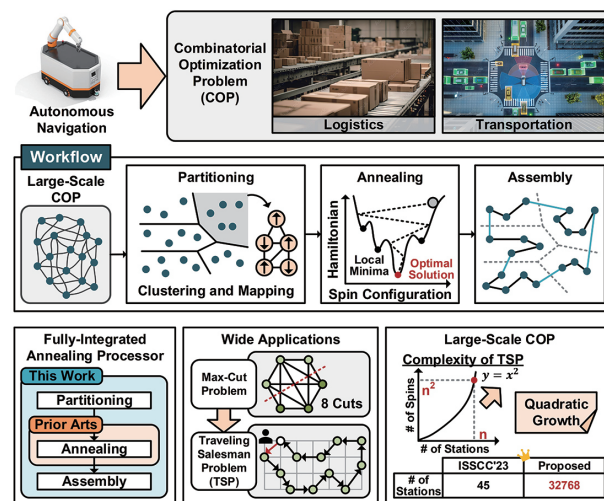
AI晶片設計、基頻通訊積體電路、生醫訊號處理器等高能效客製化晶片

◆ 作品摘要 ◆

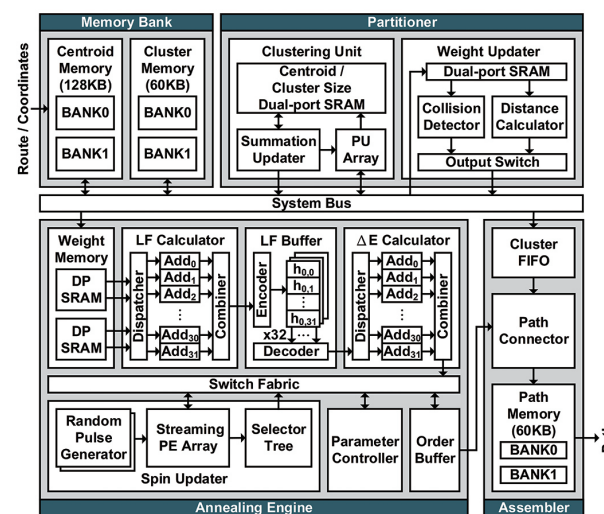
自動導航系統為基於事先標示之特徵自動控制機器人並尋找前進方向之系統，可用於工業生產、無人駕駛等多領域。其常見的應用情境為路徑規劃與避免運載機器人間的碰撞。然而，以上的應用都屬於高複雜度的組合最佳化問題，沒有唯一解且非常耗時。近年來，數位退火處理器晶片透過硬體加速運算，首先將組合最佳化問題對應到易辛模型，並運行平行化模擬退火以找到最佳解。然而過去文獻所提出之數位退火處理器，僅能支援理論性的最佳化問題，像是最大切割問題、布林可滿足性問題，抑或是被問題的規模所限制，此外過去文獻不支援問題映射（將最佳化問題對應至易辛模型），而該步驟為數位退火求解之重要的前處理步驟。本研究提出應用於大規模自動導航最佳化之量子啟發式數位退火處理器晶片，可即時規畫路徑與避免運載機器人碰撞。其運作流程包含問題分割（包含分群以及問題映射）、數位退火、問題再組合（問題反映射以及遍歷所有解）。

本作品在硬體面積與延遲最佳化部分，透過資料正規化（18位元降至5位元），降低退火模組面積37%，並維持問題解之精確性。本團隊設計自旋粒子更新模組，支援不同拓樸圖；透過兩階段選擇器，降低85%之關鍵路徑延遲。此研究支援多晶片串接以提升運算平行度，晶片使用兩時域之時鐘，同時滿足晶片運算之高效率與跨晶片傳輸之時序規格。

本研究運用演算法及硬體共同最佳化技術，在能量與面積效率比起先進CPU提升四個數量級，實現低功耗且精準即時找到最佳化問題的解，整體系統功耗為110mW，與過去退火處理器相比，此晶片降低3.8倍的功耗並提升了28.6倍正規化粒子自旋之面積效率。



圖一 數位退火處理器流程圖與其設計挑戰。



圖二 系統架構圖。

◆ Abstract ◆

Autonomous navigation has become a critical technology widely utilized in sectors such as logistics and transportation, significantly enhancing efficiency and safety. The widespread implementation of automatic guided vehicles in industrial environments underscores the significant adoption of this technology. These applications primarily focus on two critical functions: path planning and collision avoidance, both of which are categorized as combinatorial optimization problems (COPs). Solving these problems optimally is a complex and time-intensive process. Solving COPs optimally involves mapping them onto an Ising model, characterized by spin couplings, and employing annealing to iteratively refine solutions until reaching the system's lowest energy state, the Hamiltonian. Recent advancements in digital annealers have accelerated the solving of these complex COPs. However, limitations exist in current designs, with some restricted to smaller-scale problems due to limited spins, and others capable of addressing larger issues but constrained by sparse spin connections. The necessity of mapping COPs onto an Ising model and the reverse process remains an unresolved challenge for existing digital annealers.

This work introduces a fully integrated annealing processor designed for optimizing large-scale COPs. The proposed workflow includes partitioning (clustering and mapping), annealing, and assembly (de-mapping and solution traversal). By adopting the clustering approach, a large-scale problem is divided into subproblems, expanding the number of stations in the traveling salesman problem (TSP) by a factor of 1024x, and then mapped onto the Ising model. Subsequently, annealing is executed on each subproblem to find their optimal solutions. These solutions are then de-mapped back onto the original COP, deriving an overall solution through the traversal of all subproblem solutions. This work presents the first silicon implementation of energy-efficient hardware mapping for autonomous navigation optimization, which includes path planning and collision avoidance.

The system architecture of the annealing processor consists of a Partitioner, an Annealing Engine, and an Assembler. The energy and area efficiencies are enhanced through algorithm-architecture optimization. The problem mapping is simplified to reduce computational complexity and memory requirements. Efficient weight normalization is conducted to lower the area further. A combination of an adaptable spin update scheme and operator sharing for the annealing process serves to reduce the latencies and area, respectively. In the Annealing Engine, the Streaming PE Array is designed to accelerate the exploration of feasible flips through parallel processing. Meanwhile, the Selector Tree employs a two-stage tree architecture to optimize the critical path.

Fabricated in a 40-nm CMOS technology, the chip integrates 3.9M logic gates in a 4.56 mm² area. Compared to previous works, this chip can address complete COPs of the largest scale (with up to 32,768 stations), surpassing the capabilities of prior arts (which supported up to 45 stations). The chip achieves a power consumption of 0.3μW/spin within a silicon area of 516μm²/spin. This represents a reduction in spin power by 3.8x and a decrease in spin area by 28.6x, compared to previous works.

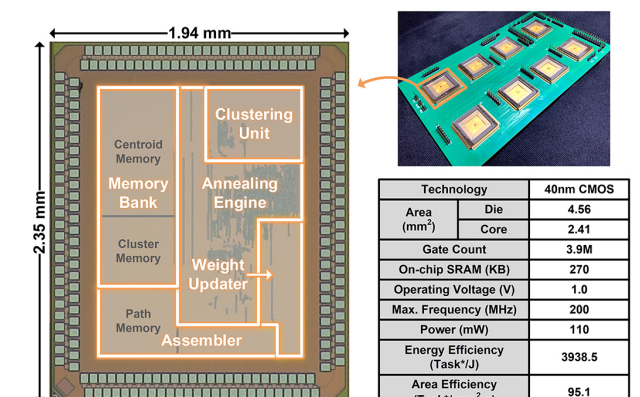


Fig. 3 Chip micrograph.

*Task: one 101-station path planning problem